Current Biology, Vol. 13, R606–R607, August 5, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0960-9822(03)00527-X

Phage Evolution: New Worlds of Genomic Diversity

Dispatch

R. Thane Papke and W. Ford Doolittle

A recent comparative survey of genomes of phages infecting mycobacteria reveals a vast combinatorial network of gene rearrangements and may provide general models for pattern and process in genome evolution.

In this still-genomic era — post-genomic only for human geneticists — the genomes of more and more closely related organisms are being sequenced, and this for many reasons. In GenBank, as of the writing of this dispatch, there exist sequences for six *Escherichia coli* genomes, seven *Staphylococcus aureus* genomes, nine *Mycobacterium* sp. genomes and seven *Bacillus* sp. genomes, among many others. The major focus of prokaryotic genomic sequencing has been on pathogens. To truly appreciate the diversity of microorganisms and the mechanisms behind such diversity, we need to move towards non-pathogenic organisms, and we need to realize that much of microbial genomic diversity resides in the genomes of the viruses which infect them.

Pedulla *et al*. [1] recently reported the sequencing of ten genomes of mycobacteriophages (in sum nearly 1 Mbp), comparing them to each other and to four previously sequenced mycobacteriophage genomes. Mycobacteriophage infect mycobacteria, the most notorious of which cause diseases such as tuberculosis and leprosy, although many grow harmlessly in soil. A total of 1659 open reading frames were identified among the 14 mycobacteriophage genomes, of which approximately 50% had no matches in queried data bases. An additional 37.5% of the identified open reading frames have only been found in mycobacteriophage, meaning that nearly 88% of the total open reading frames identified in the study are unique to the mycobacteriophage gene pool. This number is bound to decrease as more phage genetic diversity is surveyed, but, as the authors state [1], their results do suggest that there is vast unexplored genetic diversity — many as yet unknown genes and activities — locked up in bacteriophage. Indeed, they venture that "bacteriophages perhaps represent the biggest unexplored reservoir of sequence information in the biosphere".

Wild optimism regarding tremendous phage diversity comes from other sources as well: observations in the marine and thermal environments have been especially fruitful. For instance, genetic diversity of uncultured marine bacterial viruses (isolated as particles) from two different samples demonstrated that 65–75% of the cloned DNA had no significant matches in GenBank (E value of <0.001) [2]. From acidic hot springs, several notably different phage morphologies have been isolated from *Sulfolobus* sp. — which grows optimally at 80°C and pH 3 — including shapes resembling spindles, helical rods, flexible rods with attachment fibers, bearded droplets, icosahedral shapes with mushroom-like projections at the vertices and spindles with appendages, each of which may represent new virus families [3]. As well, if there is as much genetic diversity among different isolated *Sulfolobus* viruses with the same morphology as has been witnessed in the morphologically similar mycobacteriophages described in [1], then global phage diversity, in terms of novel genes and novel combinations of genes, must indeed be vast.

It is inescapable that lateral gene transfer has played a role in the evolution of the examined phages. In addition to the mosaic genomic structure that will be discussed further below, Pedulla *et al*. [1] found that up to 13% of the total open reading frames they defined are homologous to open reading frames from non-mycobacteriophage species, such as Bacteria. Furthermore, they argue that the occurrence of lateral gene transfer among the mycobacteriophage populations is even more frequent than that apparently occurring between mycobacteriophage and their hosts or other bacteriophage.

The evidence here is that approximately 90% of the identifiable homologues are shared by one or more of the sequenced mycobacteriophage genomes, although none is common to all of them. In pairwise comparisons of mycobacteriophage genomes, every phage shared at least one gene with another phage, and a majority of pairs appeared to have five or more genes in common. In several pairs a significant portion of the total genome was composed of homologous genes. No doubt, much of this gene sharing reflects homologous recombination, which is known to be more frequent between sequences that are more similar from studies on *Bacillus* sp., enterobacteria and yeast [4–6].

Much of the gene sharing and exchange process revealed by Pedulla *et al*. [1], however, affects unrelated genes, which themselves have little sequence similarity. Perhaps one of the more interesting aspects of the Padulla *et al*. [1] study is that the mass sequencing of many mycobacteriophage genomes provides many lines of evidence favoring a new and interesting theory proposed by Hendrix [7] regarding the evolution of mycobacteriophages and perhaps all tailed bacteriophages.

The mosaic nature of phage genomes has been known for some time [8], but the reason for this mosaicism has been uncertain. It was originally suggested [9] that site-specific 'linker' sequences are responsible for integrating newly acquired nonhomologous DNA in a nonrandom fashion, for example, exactly between genes; but very little evidence has been found supporting the putative linker sequences. Hendrix [7] proposed that illegitimate recombination

Department of Biochemistry, Canadian Institute for Advanced Research, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.

(by whatever means) occurs arbitrarily throughout the phage genome, and that the nonrandom appearance of recombination events is the result of a natural selection gauntlet that eliminates 'offspring' that have DNA inserted in the middle of useful relevant genes. This seems to be an extremely inefficient method for generating diversity, given the number of possible offspring from a single infection and the possibility that no offspring are viable. But as Darwin knew, natural selection through generation of random variants is not about efficiency, and the data of Padulla *et al*. [1] support Hendrix's [7] interpretation.

This new paper [1] about phages has, we think, three important lessons for those of us in the larger community of researchers interested in genomic diversity of their prokaryotic hosts. First, that phages provide a vast community of exchangeable genes which overlaps that of their hosts, and may drive its evolution. Second, collections of phage genomes, such as the one described here, may provide invaluable data sets for building theoretical models which see evolution's *pattern* as a web, not a tree, and seek to measure the relative importance of lateral versus vertical inheritance. Third, these phages may also be telling us something important about the *process* of genome evolution through exchange. Maybe no special recombination mechanisms need be involved: the gene transfer we see may be very rare survivors — because they function and provide value — of the ruthless selective winnowing of vastly many accidental illegitimate events.

References

1. Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., *et al*. (2003). Origins of highly mosaic mycobacteriophage genomes. Cell *113*, 171–182.
2. Breitbart, M., Salamon, P. Andresen,B., Mahaffy, J.M., Segall, A.M., Mead, d., Azam, F. and Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U.S.A. *99*, 14250–14255.
3. Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T. and Young, M.J. (2001). Viruses from extreme thermal environments. Proc. Natl. Acad. Sci. U.S.A. *98*, 13341–13345.
4. Roberts, M. and Cohan, F. (1993). The effect of DNA seqence divergence on sexual isolation in *Bacillus.* Genetics *134*, 401–408.
5. Vulic, M., Dionisio, F., Taddei, F. and Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. U.S.A. *94*, 9763–9767.
6. Datta, A., Hendrix, M., Lipsitch, M. and Jinks-Robertson, S. (1997). Duel roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in Yeast. Proc. Natl. Acad. Sci. U.S.A. *94*, 9757–9762.
7. Hendrix, R.W. (2000). Bacteriophages: evolution of the majority. Theor. Popul. Biol. *61*, 471–480.
8. Simon, M.N., Davis, R.W. and Davidson, N. (1971). Heteroduplexes of DNA molecules of lambdoid phages: physical mapping of their base sequence relationships by electron microscopy. In The Bacteriophage Lambda, A.D. Hershey, ed. (Cold Spring Harbor laboratory, Cold Spring Harbor, N.Y.), pp. 313–328.
9. Susskind, M.M. and Botstein, D. (1978). Molecular genetics of bacteriophage P22. Microbiol. Rev. *42*, 385–413.